Europäisches Patentamt
European Patent Office
Office européen des brevets

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02021916.8

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 02021916.8
Demande no:

Anmeldetag:
Date of filing: 27.09.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

DeveloGen Aktiengesellschaft für
entwicklungsbiologische Forschung
Rudolf-Wissell-Strasse 28
37079 Göttingen
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

SKRP, astray, string VACM associated with metabolic control

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

A61K38/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

WEICKMANN & WEICKMANN

Patentanwälte
European Patent Attorneys · European Trademark Attorneys

DIPL.-ING. H. WEICKMANN (bis 31.1.01)
DIPL.-ING. F. A. WEICKMANN
DIPL.-CHEM. B. HUBER
DR.-ING. H. LISKA
DIPL.-PHYS. DR. J. PRECHTEL
DIPL.-CHEM. DR. B. BÖHM
DIPL.-CHEM. DR. W. WEISS
DIPL.-PHYS. DR. J. TIESMEYER
DIPL.-PHYS. DR. M. HERZOG
DIPL.-PHYS. B. RUTTENSPERGER
DIPL.-PHYS. DR.-ING. V. JORDAN
DIPL.-CHEM. DR. M. DEY
DIPL.-FORSTW. DR. J. LACHNIT

Unser Zeichen:
29022P EP/WWld

Anmelder:
DeveloGen Aktiengesellschaft für
entwicklungsbiologische Forschung
Rudolf-Wissell-Straße 28

37079 Göttingen

SKRP, astray, string, VACM associated with metabolic control

SKRP, astray, string, VACM associated with metabolic control

Description

This invention relates to the use of nucleic acid sequences encoding CG7042, astray, string, or CG1401 homologous proteins, and the polypeptides encoded thereby and to the use thereof in the diagnosis, study, prevention, and treatment of diseases and disorders related to body-weight regulation, for example, but not limited to, metabolic diseases or dysfunctions such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea.

There are several metabolic diseases of human and animal metabolism, eg., obesity and severe weight loss, that relate to energy imbalance where caloric intake versus energy expenditure is imbalanced. Obesity is one of the most prevalent metabolic disorders in the world. It is still a poorly understood human disease that becomes as a major health problem more and more relevant for western society. Obesity is defined as a body weight more than 20% in excess of the ideal body weight, frequently resulting in a significant impairment of health. It is associated with an increased risk for cardiovascular disease, hypertension, diabetes, hyperlipidaemia and an increased mortality rate. Besides severe risks of illness, individuals suffering from obesity are often isolated socially.

Obesity is influenced by genetic, metabolic, biochemical, psychological, and behavioral factors. As such, it is a complex disorder that must be addressed on several fronts to achieve lasting positive clinical outcome. Since obesity is not to be considered as a single disorder but as a heterogeneous group of conditions with (potential) multiple causes, it is

also characterized by elevated fasting plasma insulin and an exaggerated insulin response to oral glucose intake (Koltermann, J. Clin. Invest 65, 1980, 1272-1284). A clear involvement of obesity in type 2 diabetes mellitus can be confirmed (Kopelman, Nature 404, 2000, 635-643).

The molecular factors regulating food intake and body weight balance are incompletely understood. Even if several candidate genes have been described which are supposed to influence the homeostatic system(s) that regulate body mass/weight, like leptin, VCPI, VCPL or the peroxisome proliferator-activated receptor-gamma co-activator, the distinct molecular mechanisms and/or molecules influencing obesity or body weight/body mass regulations are not known. In addition, several single-gene mutations resulting in obesity have been described in mice, implicating genetic factors in the etiology of obesity. (Friedman and Leibel, 1990, Cell 69: 217-220). In the ob mouse a single gene mutation (obese) results in profound obesity, which is accompanied by diabetes (Friedman et. al., 1991, Genomics 11: 1054-1062).

Therefore, the technical problem underlying the present invention was to provide for means and methods for modulating (pathological) metabolic conditions influencing body-weight regulation and/or energy homeostatic circuits. The solution to said technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the present invention relates to genes with novel functions in body-weight regulation, energy homeostasis, metabolism, and obesity. The present invention discloses specific genes involved in the regulation of body-weight, energy homeostasis, metabolism, and obesity, and thus in disorders related thereto such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea. In particular, the present invention

describes the human CG7042, astray, string, or CG1401 homologous genes as being involved in those conditions mentioned above.

The term 'GenBank Accession number' relates to NCBI GenBank database entries (Benson et al, Nucleic Acids Res. 28, 2000, 15-18).

Stress-activated protein kinase (SAPK) pathway-regulating phosphatase 1 (SKRP1) is a member of the mitogen-activated protein kinase (MAPK) phosphatase (MKP) family. SKRP1 interacts physically with the MAPK kinase MKK7, a c-Jun N-terminal kinase (JNK) activator, and inactivates the MAPK JNK pathway. SKRP1 contributes to the precise regulation of JNK signaling and plays a scaffold role for the JNK signaling (Zama T. et al., (2002) J Biol Chem 277(26):23919-23926). Mitogen-activated protein kinases (MAPKs) are activated in response to various extracellular stimuli, and their activities are regulated by upstream activating kinases and protein phosphatases such as MAPK phosphatases (MKPs). SKRP1, a member of the MKP family, contains an extended active site sequence motif conserved in all MKPs but lacks a Cdc25 homology domain. SKRP1 interacts with its physiological substrate JNK through MKK7, thereby leading to the precise regulation of JNK activity in vivo (Zama T. et al., (2002) J Biol 277(26):23909-23918).

Another dual specifity protein phosphatase and member of the MKP family, MAPK phosphatase-1 (MKP-1), has been studied in diabetic rats. Protein expression of MKP-1, a dual specificity phosphatase that inactivates MAPK, was decreased in streptozotocin-induced diabetes mellitus (DM) rats. Glomerular MAPK is activated in DM by multiple mechanisms i.e., increases in protein contents, increased phosphorylation, and decreased dephosphorylation of the enzyme due to suppression of MKP-1. These alterations may have an implication in the pathogenesis of diabetic nephropathy (Awazu M. et al., (1999) J Am Soc Nephrol 10(4):738-745). Gene expression of MKP-1 in hepatectomized liver in type 1 diabetic BB

rats is changed (Chin S. et al., (1995) Am J Physiol 269(4 Pt 1):E691-700).

Phosphoserine phosphatase (PSP) is a member of a large class of enzymes that catalyze phosphoester hydrolysis using a phosphoaspartate-enzyme intermediate. PSP is a likely regulator of the steady-state d-serine level in the brain, which is a critical co-agonist of the N-methyl-d-aspartate type of glutamate receptors (Wang W. et al., (2002) J Mol Biol 319(2):421-431). PSP belongs to a class of phosphotransferases forming an acylphosphate during catalysis (Collet J. F. et al., (1999) J Biol Chem 274(48):33985-33990). An induction of diabetes with streptozotocin resulted in significant increases in GLUT-4 phosphorylation. In contrast to normal cells, insulin failed to promote GLUT-4 recruitment to the plasma membranes and its dephosphorylation in diabetic adipocytes. At the same time, diabetes appears to induce redistribution of PSP, resulting in lower cytosolic activity and higher particulate activity. It also appears that the existence of highly phosphorylated GLUT-4 in the plasma membranes of diabetic adipocytes resulted from its inability to interact with particulate PSP (Begum N. and Draznin B., (1992) J Clin Invest 90(4):1254-1262). Calcium-induced and cAMP-mediated phosphorylation and activation of inhibitor 1 results in inhibition of PSPase activity in insulin target cells. The inhibition of PSP may cause inappropriate serine dephosphorylation of substrates of insulin action resulting in insulin resistance (Begum N. et al., (1992) J Biol Chem 267(9):5959-5963).

String (stg) is required for mitosis early in development and is transcribed in a dynamic pattern that anticipates the pattern of embryonic cell divisions. Regulated expression of stg mRNA controls the timing and location of zygotically driven embryonic cell divisions (Edgar B. A. and O'Farrell P. H., (1989) Cell 57: 177-187; Edgar B. A. and O'Farrell P. H., (1990) Cell 62: 469-480). stg regulation is a critical part of the control of early entry into mitosis in some, but not all, G2-arrested imaginal cells. stg

is essential for the generation of the adult cuticle (Kylsten P. and Saint R., (1997) Dev Biol. 192(2): 509-522). stg is required for completion of daughter centriole assembly in embryos (Vidwans S. J. et al., (1999) J Cell Biol 147(7):1371-1378).

The Cdc25 family of protein phosphatases positively regulates the cell division cycle by activating cyclin-dependent protein kinases. In humans and rodents, three Cdc25 family members denoted Cdc25A, -B, and -C have been identified. The murine forms of Cdc25 exhibit distinct patterns of expression both during development and in adult mouse tissues. Mice lacking Cdc25C (Cdc25C(-/-) mice) are viable and do not display any obvious abnormalities. Cdc25C is expressed most abundant in testis, followed by thymus, ovary, spleen, and intestine. Cdc25A and/or Cdc25B may compensate for loss of Cdc25C in the mouse (Chen M. S. et al. (2001) Mol Cell Biol 21(12):3853-3861). Cdc25 phosphatases, which dephosphorylate cyclin-dependent kinases, are overexpressed in many human tumors (Pestell K. E. et al., (2000) Oncogene 19(56):6607-6612).

Vasopressin-activated $Ca^{(2+)}$-mobilizing (VACM-1), a cullin gene family member, regulates cellular signaling. The VACM-1 receptor binds arginine vasopressin (AVP) but does not have amino acid sequence homology with the traditional AVP receptors. VACM-1, however, is homologous with a cullin family of proteins that has been implicated in the regulation of cell cycle through the ubiquitin-mediated degradation of cyclin-dependent kinase inhibitors. The effects of VACM-1 expression on the $Ca^{(2+)}$ and cAMP-dependent signaling pathway were examined. Expression of the VACM-1 gene reduced cAMP production (Burnatowska-Hledin M. et al., (2000) Am J Physiol Cell Physiol 279(1):C266-273).

So far, it has not been described that the CG7042, astray, string, or CG1401 proteins of the invention and homologous proteins are involved in the regulation of energy homeostasis and body-weight regulation and

related disorders, and thus, no functions in metabolic diseases and dysfunctions and other diseases as listed above have been discussed.

In this invention, we demonstrate that the correct gene dose of CG7042, astray, string, or CG1401 is essential for maintenance of energy homeostasis. The fly Drosophila melanogaster was used as model organism for the identification of proteins involved in the energy homeostasis. Drosophila melanogaster is one of the most intensively studied organisms in biology and serves as a model system for the investigation of many developmental and cellular processes common to higher eukaryotes, including humans (see, for example, Adams et al., Science 287: 2185-2195 (2000)). The success of Drosophila melanogaster as a model organism is largely due to the power of forward genetic screens to identify the genes that are involved in a biological process (see, Johnston Nat Rev Genet 3: 176-188 (2002); Rorth, Proc Natl Acad Sci U S A 93: 12418-12422 (1996)). In this invention, we have used a genetic screen to identify, that mutations of CG7042, astray, string, or CG1401 homologous genes cause changes in the body weight which is reflected by a significant change in the triglyceride levels. Triglycerides are the most efficient storage for energy in cells, and are significantly increased in obese patients.

In this invention the terms CG7042, astray, string, CG1401, or CG7042 proteins and nucleic acids, include Drosophila and mammalian, preferably human, homolog polypeptides or proteins and nucleic acid sequences encoding those proteins, particularly stress-activated protein kinase pathway-regulating phosphatase (SKRP), phosphoserine phosphatase (PSP), cell division cycle 25 (CDC25) proteins, or cullin (VACM) proteins and nucleic acid sequences encoding those proteins.

The present invention discloses that CG7042, astray, string, or CG1401 homologous proteins are regulating the energy homeostasis and fat

metabolism especially the metabolism and storage of triglycerides, and polynucleotides, which identify and encode the proteins disclosed in this invention. The invention also relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides of the invention. The invention also relates to the use of these polynucleotides, polypeptides and effectors thereof, e.g. antibodies, anti-sense molecules, ribozymes, aptamers, low-molecular weight molecules etc., in the diagnosis, study, prevention, and treatment of diseases and disorders, for example, but not limited to, metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea.

CG7042, astray, string, or CG1401 homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are nucleic acids encoding the human CG7042, astray, string, or CG1401 homologs (in particular the human stress-activated protein kinase pathway-regulating phosphatase 1 (SKRP1), phosphoserine phosphatase (PSP), cell division cycle 25A, B, and C (CDC25A, CDC25B, CDC25C) proteins, or cullin 5 (VACM-1) proteins and the protein similar to stress-activated protein kinase pathway-regulating phosphatase (SKRP), phosphoserine phosphatase (PSP), cell division cycle 25 (CDC25) proteins, or cullin (VACM) proteins).

The invention particularly relates to a nucleic acid molecule encoding a polypeptide contributing to regulating the energy homeostasis and the metabolism of triglycerides, wherein said nucleic acid molecule comprises

(a) the nucleotide sequence of Drosophila CG7042, astray, string, or CG1401, human CG7042, astray, string, or CG1401 homologs, and/or a sequence complementary thereto,

(b) a nucleotide sequence which hybridizes at 50°C in a solution containing 1 x SSC and 0.1% SDS to a sequence of (a),

(c) a sequence corresponding to the sequences of (a) or (b) within the degeneration of the genetic code,

(d) a sequence which encodes a polypeptide which is at least 85%, preferably at least 90%, more preferably at least 95%, more preferably at least 98% and up to 99,6% identical to the amino acid sequences of the CG7042, astray, string, or CG1401 protein, preferably of the human CG7042, astray, string, or CG1401 homologs,

(e) a sequence which differs from the nucleic acid molecule of (a) to (d) by mutation and wherein said mutation causes an alteration, deletion, duplication and/or premature stop in the encoded polypeptide or

(f) a partial sequence of any of the nucleotide sequences of (a) to (e) having a length of at least 15 bases, preferably at least 20 bases, more preferably at least 25 bases and most preferably at least 50 bases.

The present invention relates to genes with novel functions in body-weight regulation, energy homeostasis, metabolism, and obesity, fragments of said genes, polypeptides encoded by said genes or fragments thereof, and effectors e.g. antibodies, biologically active nucleic acids, such as antisense molecules or ribozymes, aptamers, peptides or low-molecular weight organic compounds recognizing said polynucleotides or polypeptides.

The invention is based on the finding that CG7042, astray, string, or CG1401 and the polynucleotides encoding these, are involved in the regulation of triglyceride storage and therefore energy homeostasis. To find genes with novel functions in energy homeostasis, metabolism, and obesity, a functional genetic screen was performed with the model

organism Drosophila melanogaster (Meigen). One resource for screening was a Drosophila melanogaster stock collection of EP-lines. The P-vector of this collection has Gal4-UAS-binding sites fused to a basal promoter that can transcribe adjacent genomic Drosophila sequences upon binding of Gal4 to UAS-sites (Brand A. H. and Perrimon N., (1993) Development 118:401-415; Rorth P., (1996) Proc Natl Acad Sci U S A 93:12418-12422). This enables the EP-line collection for overexpression of endogenous flanking gene sequences. In addition, without activation of the UAS-sites, integration of the EP-element into the gene is likely to cause a reduction of gene activity, and allows determining its function by evaluating the loss-of-function phenotype.

Triglycerides are the most efficient storage for energy in cells, and obese patients mainly show a significant increase in the content of triglycerides. In order to isolate genes with a function in energy homeostasis, several thousand EP-lines were tested for their triglyceride content after a prolonged feeding period (see Examples for more detail). Lines with significantly changed triglyceride content were selected as positive candidates for further analysis. The change of triglyceride content due to the loss of a gene function suggests gene activities in energy homeostasis in a dose dependent manner that controls the amount of energy stored as triglycerides.

In this invention, the content of triglycerides of a pool of flies with the same genotype after feeding for six days was analyzed using a triglyceride assay. Male flies homozygous for the integration of vectors for Drosophila lines HD-EP(3)37139, HD-EP(3)36956, HD-EP(3)36964, HD-EP(3)36936, and HD-EP(3)36858 were analyzed in an assay measuring the triglyceride contents of these flies, illustrated in more detail in the EXAMPLES section. The results of the triglyceride content analysis are shown in FIGURE 1, 3, 5, and 7, respectively.

Genomic DNA sequences were isolated that are localized to the EP vector (herein HD-EP(3)37139, HD-EP(3)36956, HD-EP(3)36964, HD-EP(3)36936, and HD-EP(3)36858) integration. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly; see also FlyBase (1999) Nucleic Acids Research 27:85-88) were screened thereby identifying the integration site of the vectors, and the corresponding gene, described in more detail in the EXAMPLES section. The molecular organization of the gene is shown in FIGURE 2, 4, 6, and 8, respectively.

The present invention further describes polypeptides comprising the amino acid sequences of the proteins of the invention and homologous proteins. Based upon homology, the proteins of the invention and each homologous protein or peptide may share at least some activity.

The invention also encompasses polynucleotides that encode the proteins of the invention and homologous proteins. Accordingly, any nucleic acid sequence, which encodes the amino acid sequences of the proteins of the invention and homologous proteins, can be used to generate recombinant molecules that express the proteins of the invention and homologous proteins. In a particular embodiment, the invention encompasses a nucleic acid encoding Drosophila CG7042, astray, string, or CG1401, or human CG7042, astray, string, or CG1401 homologs; referred to herein as the proteins of the invention. It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding the proteins, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. The invention contemplates each and every possible variation of nucleotide sequence that can be made by selecting combinations based on possible codon choices.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in

particular, those of the polynucleotide encoding the proteins of the invention, under various conditions of stringency. Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex or probe, as taught in Wahl, G. M. and S. L. Berger (1987: Methods Enzymol. 152:399-407) and Kimmel, A. R. (1987; Methods Enzymol. 152:507-511), and may be used at a defined stringency. Preferably, hybridization under stringent conditions means that after washing for 1 h with 1 x SSC and 0.1% SDS at 50°C, preferably at 55°C, more preferably at 62°C and most preferably at 68°C, particularly for 1 h in 0.2 x SSC and 0.1% SDS at 50°C, preferably at 55°C, more preferably at 62°C and most preferably at 68°C, a positive hybridization signal is observed. Altered nucleic acid sequences encoding the proteins which are encompassed by the invention include deletions, insertions or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent protein.

The encoded proteins may also contain deletions; insertions or substitutions of amino acid residues, which produce a silent change and result in functionally equivalent proteins. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological activity of the protein is retained. Furthermore, the invention relates to peptide fragments of the proteins or derivatives thereof such as cyclic peptides, retro-inverso peptides of peptide mimetics having a length of at least 4, preferably at least 6 and up to 50 amino acids.

Also included within the scope of the present invention are alleles of the genes encoding the proteins of the invention and homologous proteins. As used herein, an 'allele' or 'allelic sequence' is an alternative form of the gene, which may result from at least one mutation in the nucleic acid sequence. Alleles may result in altered mRNAs or polypeptides whose

structures or function may or may not be altered. Any given gene may have none, one or many allelic forms. Common mutational changes, which give rise to alleles, are generally ascribed to natural deletions, additions or substitutions of nucleotides. Each of these types of changes may occur alone or in combination with the others, one or more times in a given sequence.

The nucleic acid sequences encoding the proteins of the invention and homologous proteins may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements.

In order to express a biologically active protein, the nucleotide sequences encoding the proteins or functional equivalents, may be inserted into appropriate expression vectors, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods, which are well known to those skilled in the art, may be used to construct expression vectors containing sequences encoding the proteins and the appropriate transcriptional and translational control elements. Regulatory elements include for example a promoter, an initiation codon, a stop codon, a mRNA stability regulatory element, and a polyadenylation signal. Expression of a polynucleotide can be assured by (i) constitutive promoters such as the Cytomegalovirus (CMV) promoter/enhancer region, (ii) tissue specific promoters such as the insulin promoter (see, Soria et al., 2000, Diabetes 49:157), SOX2 gene promotor (see Li et al., 1998, Curr. Biol. 8:971-4), Msi-1 promotor (see Sakakibara et al., 1997, J. Neuroscience 17:8300-8312), alpha-cardia myosin heavy chain promotor or human atrial natriuretic factor promotor (Klug et al., 1996, J. clin. Invest 98:216-24; Wu et al., 1989, J. Biol. Chem. 264:6472-79)or (iii) inducible promoters such as the tetracycline inducible system. Expression vectors can also contain a selection agent or marker gene that confers antibiotic resistance such as the neomycin, hygromycin

or puromycin resistance genes. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y. and Ausubel, F.M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.

In a further embodiment of the invention, natural, modified or recombinant nucleic acid sequences encoding the proteins of the invention and homologous proteins may be ligated to a heterologous sequence to encode a fusion protein.

A variety of expression vector/host systems may be utilized to contain and express sequences encoding the proteins or fusion proteins. These include, but are not limited to, micro-organisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus, adenovirus, adeno-associated virus, lentiverus, retrovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or PBR322 plasmids); or animal cell systems.

The presence of polynucleotide sequences of the invention in a sample can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or portions or fragments of said polynucleotides. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences specific for the gene to detect transformants containing DNA or RNA encoding the corresponding protein. As used herein 'oligonucleotides' or 'oligomers' refer to a nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides,

preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides, which can be used as a probe or amplimer.

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting polynucleotide sequences include oligo-labeling, nick translation, end-labeling of labeled RNA probes, PCR amplification using a labeled nucleotide, or enzymatic synthesis. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, Mich.); Promega (Madison Wis.); and U.S. Biochemical Corp., (Cleveland, Ohio).

The presence of proteins of the invention in a sample can be determined by immunological methods or activity measurement. A variety of protocols for detecting and measuring the expression of proteins, using either polyclonal or monoclonal antibodies specific for the protein or reagents for determining protein activity are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on the protein is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, Minn.) and Maddox, D. E. et al. (1983; J. Exp. Med. 158:1211-1216).

Suitable reporter molecules or labels, which may be used, include radionuclides, enzymes, fluorescent, chemiluminescent or chromogenic agents as well as substrates, co-factors, inhibitors, magnetic particles, and the like.

The nucleic acids encoding the proteins of the invention can be used to generate transgenic animal or site specific gene modifications in cell lines. Transgenic animals may be made through homologous recombination, where the normal locus of the genes encoding the proteins of the invention is altered. Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include plasmids, retrovirusses and other animal virusses, YACs, and the like. The modified cells or animal are useful in the study of the function and regulation of the proteins of the invention. For example, a series of small deletions and/or substitutions may be made in the genes that encode the proteins of the invention to determine the role of particular domains of the protein, functions in pancreatic differentiation, etc.

Specific constructs of interest include anti-sense molecules, which will block the expression of the proteins of the invention, or expression of dominant negative mutations. A detectable marker, such as for example lac-Z, may be introduced in the locus of the genes of the invention, where upregulation of expression of the genes of the invention will result in an easily detected change in phenotype.

One may also provide for expression of the genes of the invention or variants thereof in cells or tissues where it is not normally expressed or at abnormal times of development. In addition, by providing expression of the proteins of the invention in cells in which they are not normally produced, one can induce changes in cell behavior.

DNA constructs for homologous recombination will comprise at least portions of the genes of the invention with the desired genetic modification, and will include regions of homology to the target locus. DNA constructs for random integration need not include regions of homology to mediate recombination. Conveniently, markers for positive and/or negative selection are included. Methods for generating cells having targeted gene

modifications through homologous recombination are known in the art. For embryonic stem (ES) cells, an ES cell line may be employed, or embryonic cells may be obtained freshly from a host, e.g. mouse, rat, guinea pig etc. Such cells are grown on an appropriate fibroblast-feeder layer or grown in presence of leukemia inhibiting factor (LIF).

When ES or embryonic cells or somatic pluripotent stem cells have been transformed, they may be used to produce transgenic animals. After transformation, the cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be detected by employing a selective medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination or integration of the construct. Those colonies that are positive may then be used for embryo manipulation and blastocyst injection. Blastocysts are obtained from 4 to 6 week old superovulated females. The ES cells are trypsinized, and the modified cells are injected into the blastocoel of the blastocyst. After injection, the blastocysts are returned to each uterine horn of pseudopregnant females. Females are then allowed to go to term and the resulting offspring screened for the construct. By providing for a different phenotype of the blastocyst and the genetically modified cells, chimeric progeny can be readily detected. The chimeric animals are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs can be maintained as allogenic or congenic grafts or transplants, or in vitro culture. The transgenic animals may be any non-human mammal, such as laboratory animal, domestic animals, etc. The transgenic animals may be used in functional studies, drug screening, etc.

Diagnostics and Therapeutics

The data disclosed in this invention show that the nucleic acids and proteins of the invention are useful in diagnostic and therapeutic applications implicated, for example but not limited to, in metabolic disorders such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea. Hence, diagnostic and therapeutic uses for the proteins of the invention nucleic acids and proteins of the invention are, for example but not limited to, the following: (i) protein therapeutic, (ii) small molecule drug target, (iii) antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) diagnostic and/or prognostic marker, (v) gene therapy (gene delivery/gene ablation), (vi) research tools, and (vii) tissue regeneration in vitro and in vivo (regeneration for all these tissues and cell types composing these tissues and cell types derived from these tissues).

The nucleic acids and proteins of the invention and effectors thereof are useful in diagnostic and therapeutic applications implicated in various applications as described below. For example, but not limited to, cDNAs encoding the proteins of the invention and particularly their human homologues may be useful in gene therapy, and the proteins of the invention and particularly their human homologues may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from, for example, but not limited to, in metabolic disorders as described above.

The nucleic acids of the invention or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acids or the proteins are to be assessed. Further antibodies that

bind immunospecifically to the novel substances of the invention may be used in therapeutic or diagnostic methods.

For example, in one aspect, antibodies, which are specific for the proteins of the invention and homologous proteins, may be used directly as an effector, e.g. an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the protein. The antibodies may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralising antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with the protein or any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. It is preferred that the peptides, fragments or oligopeptides used to induce antibodies to the protein have an amino acid sequence consisting of at least five amino acids, and more preferably at least 10 amino acids.

Monoclonal antibodies to the proteins may be prepared using any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (K hler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R. J. et al. Proc. Natl. Acad. Sci. 80:2026-2030; Cole, S. P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of 'chimeric antibodies', the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison, S. L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M. S. et al (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce single chain antibodies specific for the proteins of the invention and homologous proteins. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D. R. (1991) Proc. Natl. Acad. Sci. 88:11120-3). Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for the proteins may also be generated. For example, such fragments include, but are not limited to, the $F(ab')_2$ fragments which can be produced by Pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W. D. et al. (1989) Science 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding and immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation

between the protein and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reacive to two non-interfering protein epitopes are preferred, but a competitive binding assay may also be employed (Maddox, supra).

In another embodiment of the invention, the polynucleotides of the invention or fragments thereof or nucleic acid effector molecules such as antisense molecules or ribozymes or aptomers may be used for therapeutic purposes. In one aspect, antisense molecules may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding the proteins of the invention and homologous proteins. Thus, antisense molecules may be used to modulate protein activity or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligomers or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding the proteins. Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods, which are well known to those skilled in the art, can be used to construct recombinant vectors, which will express antisense molecules complementary to the polynucleotides of the genes encoding the proteins of the invention and homologous proteins. These techniques are described both in Sambrook et al. (supra) and in Ausubel et al. (supra). Genes encoding the proteins of the invention and homologous proteins can be turned off by transforming a cell or tissue with expression vectors, which express high levels of polynucleotides that encode the proteins of the invention and homologous proteins or fragments thereof. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by

endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, e.g. DNA, RNA or PNA, to the control regions of the genes encoding the proteins of the invention and homologous proteins, i.e., the promoters, enhancers, and introns. Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it cause inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J. E. et al. (1994) In; Huber, B. E. and B. I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y.). The antisense molecules may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples, which may be used, include engineered hammerhead motif ribozyme molecules that can be specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding the proteins of the invention and homologous proteins. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing

the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Antisense molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences. Such DNA sequences may be incorporated into a variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize antisense RNA constitutively or inducibly can be introduced into cell lines, cells or tissues. RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5′ and/or 3′ ends of the molecule or modifications in the nucleobase, sugar and/or phosphate moieties, e.g. the use of phosphorothioate or 2′ O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection and by liposome injections may be achieved using methods, which are well known in the art. Any of the therapeutic methods

described above may be applied to any suitable subject including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of the nucleic acids and the proteins of the invention and homologous nucleic acids or proteins, antibodies to the proteins of the invention and homologous proteins, mimetics, agonists, antagonists or inhibitors of the proteins of the invention and homologous proteins or nucleic acids. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone or in combination with other agents, drugs or hormones. The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries, which facilitate processing of the active compounds into preparations, which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, Pa.).

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art. For any compounds, the therapeutically effective does can be estimated initially either in cell culture assays, e.g., of preadipocyte cell lines or in animal models, usually mice, rabbits, dogs or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. A therapeutically effective dose refers to that amount of active ingredient, for example the nucleic acids or the proteins of the invention and homologous proteins or nucleic acids or fragments thereof, antibodies of the proteins of the invention and homologous proteins, which is sufficient for treating a specific condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions, which exhibit large therapeutic indices, are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage from employed, sensitivity of the patient, and the route of administration. The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors, which may be taken into account, include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency

of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week or once every two weeks depending on half-life and clearance rate of the particular formulation. Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

In another embodiment, antibodies which specifically bind to the proteins may be used for the diagnosis of conditions or diseases characterized by or associated with over- or underexpression of the proteins of the invention and homologous proteins or in assays to monitor patients being treated with the proteins of the invention and homologous proteins, or effectors thereof, e.g. agonists, antagonists, or inhibitors. Diagnostic assays include methods which utilize the antibody and a label to detect the protein in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules which are known in the art may be used several of which are described above.

A variety of protocols including ELISA, RIA, and FACS for measuring proteins are known in the art and provide a basis for diagnosing altered or abnormal levels of gene expression. Normal or standard values for gene expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibodies to the protein under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, but

preferably by photometry, means. Quantities of protein expressed in control and disease, samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides specific for the proteins of the invention and homologous proteins may be used for diagnostic purposes. The polynucleotides, which may be used, include oligonucleotide sequences, antisense RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which gene expression may be correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess gene expression, and to monitor regulation of protein levels during therapeutic intervention.

In one aspect, hybridization with probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding the proteins of the invention and homologous proteins or closely related molecules, may be used to identify nucleic acid sequences which encode the respective protein. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of the polynucleotide encoding the proteins of the invention or from a genomic sequence including promoter, enhancer elements, and introns of the naturally occurring gene. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as ^{32}P or ^{35}S or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences specific for the proteins of the invention and homologous nucleic acids may be used for the diagnosis of conditions or diseases, which are associated with the expression of the proteins. Examples of such conditions or diseases include, but are not limited to,

pancreatic diseases and disorders, including diabetes. Polynucleotide sequences specific for the proteins of the invention and homologous proteins may also be used to monitor the progress of patients receiving treatment for pancreatic diseases and disorders, including diabetes. The polynucleotide sequences may be used qualitative or quantitative assays, e.g. in Southern or Northern analysis, dot blot or other membrane-based technologies; in PCR technologies; or in dip stick, pin, ELISA or chip assays utilizing fluids or tissues from patient biopsies to detect altered gene expression.

In a particular aspect, the nucleotide sequences specific for the proteins of the invention and homologous nucleic acids may be useful in assays that detect activation or induction of various metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancers of the reproductive organs, and sleep apnea. The nucleotide sequences may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly altered from that of a comparable have hybridized with nucleotide sequences in the sample, and the presence of altered levels of nucleotide sequences encoding the proteins of the invention and homologous proteins in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials or in monitoring the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disease associated with expression of the proteins of the invention and homologous proteins, a

normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence or a fragment thereof, which is specific for the nucleic acids encoding the proteins of the invention and homologous nucleic acids, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease. Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that, which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to metabolic diseases such as described above the presence of an unusual amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the metabolic diseases and disorders.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding the proteins of the invention and homologous proteins may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically or produced from a recombinant source. Oligomers will preferably consist of two nucleotide sequences, one

with sense orientation (5'.fwdarw.3') and another with antisense (3'.rarw.5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantification of closely related DNA or RNA sequences.

In another embodiment of the invention, the nucleic acid sequences may also be used to generate hybridization probes, which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. Such techniques include FISH, FACS or artificial chromosome constructions, such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price, C. M. (1993) Blood Rev. 7:127-134, and Trask, B. J. (1991) Trends Genet. 7:149-154. FISH (as described in Verma et al. (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York, N.Y.). The results may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of the gene encoding the proteins of the invention on a physical chromosomal map and a specific disease or predisposition to a specific disease, may help to delimit the region of DNA associated with that genetic disease.

The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals. In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm

of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti, R. A. et al. (1988) Nature 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequences of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier or affected individuals.

In another embodiment of the invention, the proteins of the invention, its catalytic or immunogenic fragments or oligopeptides thereof, an in vitro model, a genetically altered cell or animal, can be used for screening libraries of compounds in any of a variety of drug screening techniques. One can identify effectors, e.g. receptors, ligands or substrates that bind to, modulate or mimic the action of one or more of the proteins of the invention. The protein or fragment thereof employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between the proteins of the invention and the agent tested, may be measured. Of particular interest are screening assays for agents that have a low toxicity for mammalian cells. The term "agent" as used herein describes any molecule, e.g. protein or pharmaceutical, with the capability of altering or mimicking the physiological function of one or more of the proteins of the invention. Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 Daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or

carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise carbocyclic or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups.

Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, nucleic acids and derivatives, structural analogs or combinations thereof. Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, etc. to produce structural analogs. Where the screening assay is a binding assay, one or more of the molecules may be joined to a label, where the label can directly or indirectly provide a detectable signal.

Another technique for drug screening, which may be used, provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to the proteins of the invention large numbers of different small test compounds, e.g. aptamers, peptides, low-molecular weight compounds etc., are provided or synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with the proteins or fragments thereof, and washed. Bound proteins are then detected by methods well known in the

art. Purified proteins can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support. In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding the protein specifically compete with a test compound for binding the protein. In this manner, the antibodies can be used to detect the presence of any peptide, which shares one or more antigenic determinants with the protein.

Finally, the invention also relates to a kit comprising at least one of

(a) a CG7042, astray, string, or CG1401 nucleic acid molecule or a fragment thereof;

(b) a vector comprising the nucleic acid of (a);

(c) a host cell comprising the nucleic acid of (a) or the vector of (b);

(d) a polypeptide encoded by the nucleic acid of (a);

(e) a fusion polypeptide encoded by the nucleic acid of (a);

(f) an antibody, an aptamer or another receptor against the nucleic acid of (a) or the polypeptide of (d) or (e) and

(g) an anti-sense oligonucleotide of the nucleic acid of (a).

The kit may be used for diagnostic or therapeutic purposes or for screening applications as described above. The kit may further contain user instructions.

The Figures show:

FIGURE 1 shows the triglyceride content of Drosophila CG7042 (GadFly Accession Number) mutants. Shown is the change of triglyceride content of HD-EP(3)37139 flies caused by integration of the P-vector into the into the annotated transcription unit (column 2) in comparison to controls containing all flies of the EP collection ('EP-control', column 1).

FIGURE 2 shows the molecular organization of the mutated CG7042 (Gadfly Accession Number) gene locus.

FIGURE 3 shows the triglyceride content of Drosophila astray (GadFly Accession Number CG3705) mutants. Shown is the change of triglyceride content of HD-EP(3)36956 and HD-EP(3)36964 flies caused by integration of the P-vector into the annotated transcription unit (columns 2 and 3, respectively) in comparison to controls containing all flies of the EP collection ('EP-control', column 1).

FIGURE 4 shows the molecular organization of the mutated astray (Gadfly Accession Number CG3705) gene locus.

FIGURE 5 shows the triglyceride content of Drosophila string (GadFly Accession Number CG1395) mutants. Shown is the change of triglyceride content of HD-EP(3)36936 flies caused by integration of the P-vector into the into the annotated transcription unit (column 2) in comparison to controls containing all flies of the EP collection ('EP-control', column 1).

FIGURE 6 shows the molecular organization of the mutated string (Gadfly Accession Number CG1395) gene locus.

FIGURE 7 shows the triglyceride content of Drosophila CG1401 (GadFly Accession Number) mutants. Shown is the change of triglyceride content of HD-EP(3)36858 flies caused by integration of the P-vector into the into the annotated transcription unit (column 2) in comparison to controls containing all flies of the EP collection ('EP-control', column 1).

FIGURE 8 shows the molecular organization of the mutated CG1401 (Gadfly Accession Number) gene locus.

The examples illustrate the invention:

Example 1: Measurement of triglyceride content

Mutant flies are obtained from a fly mutation stock collection. The flies are grown under standard conditions known to those skilled in the art. In the course of the experiment, additional feedings with bakers yeast (Saccharomyces cerevisiae) are provided for the EP-lines HD-EP(3)37139, HD-EP(3)36956, HD-EP(3)36964, HD-EP(3)36936, and HD-EP(3)36858. The average change of triglyceride content of Drosophila containing the EP-vector as homozygous viable integration was investigated in comparison to control flies (see FIGURES 1, 3, 5, and 7, respectively). For determination of triglyceride content, flies were incubated for 5 min at 90°C in an aqueous buffer using a waterbath, followed by hot extraction. After another 5 min incubation at 90°C and mild centrifugation, the triglyceride content of the flies extract was determined using Sigma Triglyceride (INT 336-10 or -20) assay by measuring changes in the optical density according to the manufacturer's protocol. As a reference the protein content of the same extract was measured using BIO-RAD DC Protein Assay according to the manufacturer's protocol. These experiments and assays were repeated several times.

The average triglyceride level of all flies of the EP collection (referred to as 'EP-control') is shown as 100% in the first columns in FIGURES 1, 3, 5, and 7. Standard deviations of the measurements are shown as thin bars.

HD-EP(3)37139 homozygous flies show constantly a higher triglyceride content than the controls (column 2 in FIGURE 1, 'HD-EP37139'). HD-EP(3)36956 and HD-EP(3)36964 homozygous flies show constantly a higher triglyceride content than the controls (column 2 in FIGURE 3,

'HD-EP36956', and column 3 in FIGURE 3 'HD-EP36964'). HD-EP(3)36936 homozygous flies show constantly a higher triglyceride content than the controls (column 2 in FIGURE 5, 'HD-EP36936'). HD-EP(3)36858 homozygous flies show constantly a higher triglyceride content than the controls (column 2 in FIGURE 7, 'HD-EP36858'). Therefore, the loss of gene activity is responsible for changes in the metabolism of the energy storage triglycerides.

Example 2: Identification of metabolic control-associated genes and proteins

(i) CG7042

Genomic DNA sequences were isolated that are localized directly adjacent to the EP vector (herein HD-EP(3)37139) integration. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly) were screened thereby confirming the homozygous viable integration site of the HD-EP(3)37139 vector into the leader sequence of cDNA CG7042-RA and into the cDNA CG7042-RB at base pair 49 in sense orientation. FIGURE 2 shows the molecular organization of this gene locus. The chromosomal localization site of integration of the vector of HD-EP(3)37139 is at gene locus 3L, 61B2. In FIGURE 2, genomic DNA sequence is represented by the assembly as a black arrow in middle of the figure that includes the integration site of HD-EP(3)37139. Ticks represent the length in basepairs of the genomic DNA (1000 base pairs per tick). Dark grey bars on the two sides, linked by dark grey lines represent cDNAs of the predicted genes (as predicted by the Berkeley Drosophila Genome Project, GadFly release 3). Predicted exons of the Drosophila cDNA CG7042 (GadFly Accession Number) are shown as dark grey bars and predicted introns as slim grey lines in the lower half of the figure. Therefore, expression of the cDNA encoding CG7042 could be affected by

integration of vectors of line HD-EP(3)37139, leading to a change in the amount of energy storage triglycerides.

(ii) astray

Genomic DNA sequences were isolated that are localized directly adjacent to the EP vectors (herein HD-EP(3)36956 and HD-EP(3)36964) integration. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly) were screened thereby confirming the homozygous viable integration site of the HD-EP(3)36956 vector 370 base pairs 5' of CG3705-RA in antisense orientation and confirming the homozygous viable integration site of the HD-EP(3)36964 vector 1003 base pairs 3' of the transcription start of CG3705-RA in antisense orientation, identified as astray (referred to as aay; GadFly Accession Number CG3705). FIGURE 4 shows the molecular organization of this gene locus. The chromosomal localization site of integration of the vectors HD-EP(3)36956 and HD-EP(3)36964 is at gene locus 3L, 67B1 (according to FlyBase), 67B4 (according to GadFly release 3). In FIGURE 4, genomic DNA sequence is represented by the assembly as a dotted grey line in the middle that includes the integration sites of vector for lines HD-EP(3)36956 and HD-EP(3)36964. Numbers represent the coordinates of the genomic DNA (starting at position 9379500 on chromosome 3L, ending at position 9382625 on chromosome 3L). The insertion sites of the P-elements in Drosophila HD-EP(3)36956 and HD-EP(3)36964 lines are shown as triangles in the "P Elements -" line and are labeled. A dark grey box on the "cDNA +" line represents the predicted gene (as predicted by the Berkeley Drosophila Genome Project, GadFly and by Magpie). Predicted exons are shown as dark grey boxes, predicted introns are shown as light grey boxes. The gene astray is labeled (referred to as aay, CG3705). Transcribed DNA sequences (ESTs) are shown as grey bars in the "EST +" line. Therefore, expression of the cDNA encoding astray could be affected

by integration of vectors of lines HD-EP(3)36956 and HD-EP(3)36964, leading to a change in the amount of energy storage triglycerides.

(iii) string

Genomic DNA sequences were isolated that are localized directly adjacent to the EP vector (herein HD-EP(3)36936) integration. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly) were screened thereby confirming the homozygous viable integration site of the HD-EP(3)36936 vector into the cDNA at base pair 144 of a Drosophila gene in sense orientation identified as string (referred to as stg; GadFly Accession Number CG1395). FIGURE 6 shows the molecular organization of this gene locus. The chromosomal localization site of integration of the vector HD-EP(3)36936 is at gene locus 3R, 98F13 (according to FlyBase), 99A5 (according to GadFly release 3). In FIGURE 6, genomic DNA sequence is represented by the assembly as a dotted grey line in the middle that includes the integration site of vector for line HD-EP(3)36936. Numbers represent the coordinates of the genomic DNA (starting at position 25065000 on chromosome 3R, ending at position 25075000 on chromosome 3R). The insertion sites of the P-element in Drosophila HD-EP(3)36936 line is shown as triangle in the "P Elements -" line and is labeled. Dark grey boxes on the "cDNA -" line, linked by light grey boxes represent the predicted genes (as predicted by the Berkeley Drosophila Genome Project, GadFly and by Magpie). Predicted exons are shown as dark grey boxes, predicted introns are shown as light grey boxes. The gene string is labeled (referred to as string, CG1395). Transcribed DNA sequences (ESTs) are shown as grey bars in the "EST -" line. Therefore, expression of the cDNA encoding string could be affected by integration of the vector of line HD-EP(3)36936, leading to a change in the amount of energy storage triglycerides.

(iv) CG1401

Genomic DNA sequences were isolated that are localized directly adjacent to the EP vector (herein HD-EP(3)36858) integration. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly) were screened thereby confirming the homozygous viable integration site of the HD-EP(3)36858 vector 1663 base pairs 5' of the cDNA of a Drosophila gene in antisense orientation, identified as CG1401-RA (referred to as GadFly Accession Number CG1401). FIGURE 8 shows the molecular organization of this gene locus. The chromosomal localization site of integration of the vector HD-EP(3)36858 is at gene locus 3R, 98F4 (according to FlyBase), 98F6 (according to GadFly release 3). In FIGURE 8, genomic DNA sequence is represented by the assembly as a dotted grey line in the middle that includes the integration site of vector for the line HD-EP(3)36858. Numbers represent the coordinates of the genomic DNA (starting at position 24873000 on chromosome 3R, ending at position 24873000 on chromosome 3R). The insertion sites of the P-elements in Drosophila HD-EP(3)36858 line is shown as box in the "P Elements +" line and is labeled. Dark grey bars on the "cDNA +" line and the "cDNA -" line, linked by light grey bars represents the predicted genes (as predicted by the Berkeley Drosophila Genome Project, GadFly and by Magpie). Predicted exons are shown as dark grey boxes, predicted introns are shown as light grey boxes. The gene CG1401 is labeled. Transcribed DNA sequences (ESTs) are shown as grey bars in the "EST +" line and the "EST -" line. Therefore, expression of the cDNA encoding CG1401 could be affected by integration of the vector of line HD-EP(3)36858, leading to a change in the amount of energy storage triglycerides.

Table 1. Molecular analysis of Drosophila CG7042, astray, string, and CG1401

Analysis	Genetic interaction
CG7042	not described
astray	not described
string	Cdc27, Cyclin E, Myb oncogene-like, Notch, Ras oncogene at 85D, noose, pannier, roughex, shotgun, armadillo, tribbles
CG1401	not described
Analysis	Protein
CG7042	protein tyrosine/serine/threonine phosphatase and/or tRNA pseudouridine synthase -> bicistronic mRNA encoding two different proteins (Flybase)
astray	phosphoserine phosphatase (Flybase)
string	protein tyrosine/serine/threonine phosphatase involved in G2/M

	transition of mitotic cell cycle which is a component of the nucleus (Flybase)
CG1401	vasopressin activated calcium mobilizing receptor (VACM; Flybase), neuropeptide receptor involved in signaling (Ca^{2+} release, cAMP signaling, interaction with PKA & PKC)
Analysis	Protein domains
CG7042	Pseudouridine synthase I (Flybase)
astray	Haloacid dehalogenase/epoxide hydrolase family, HAD-like, Membrane all-alpha, (Flybase)
string	Rhodanese/Cell cycle control phosphatase (Flybase)
CG1401	not described (Flybase)
Analysis	InterPro analysis
CG7042	CG7042-PA (203 aa): Dual specificity protein phosphatase (IPR000340), Tyrosine specific protein phosphatase and dual specificity protein phosphatase (IPR000387) CG7042-PB (306 aa): tRNA pseudouridine synthase (IPR001406)
astray	haloacid dehalogenase-like hydrolase (IPR005834)
string	M-phase inducer phosphatase (IPR000751), Rhodanese-like (IPR001763)
CG1401	Antifreeze protein, type I (IPR000104), Cullin (IPR001373)
Analysis	Locus
CG7042	3L, 61B2 (Flybase)
astray	3L, 67B1 (Flybase), 3L, 67B4 (Gadfly release 3)
string	3R, 98F13 (Flybase), 3R 99A5 (Gadfly release 3)
CG1401	3R, 98F4 (Flybase); 3R, 98F6 (Gadfly release 3)
Analysis	Ests
CG7042	CK00185, AT31323, RH73639, LD03462, GM04926, RE61580, RE01115, RE01995, RE59365
astray	RH69894, RH52615, RH73984, RH68572, RH21337, RH08743, RH04207, RH21250, RH48422, RH47146, RH56928, RH50262, RH50732, RH02523, RH07420, RH36004, RH58226, RH68424, RH33376, RH61465, RH48104, RH41563, RH09607, RH59478, RH58328, RH63182, RH58384, RH03089, RH46577, RH63664, RH64571, RH09689, RH03078, RE33309, RE15673, RE49826, RE25225, RE18081, RE13589, GM30362, LP11115, LP12306, GH22990, GH13931, GH05376, GH08849, LD06953, LD23646, AT11071, GH21532, LD38930, LD43423, GH08622, LD09949,

	GH25312, LD11335, LD04659, LD35988, GH05078, HL01504, HL05810, AT22661, AT01455, RH71749, RH55026, AT02095, LD39154, RH66377, RH59056 (Gadfly release 3)
string	LP01340, SD28250, LD12004, SD15684, SD22406, SD14743, SD14854, GM01429, SD25591, SD05666, LD43433, SD04248, SD25387, SD20377, SD19682, SD18374, LD02385, RE73447, RE72586, RE64991, RE67709, RE63711, RE66391, RE75120, RE38116, RE08989, RE56977, RE37927, RE16643, RE02486, RE48836, RE52977, RE48257, RE44539, RE50736, RE36029, LD47970, SD05639, LD47579 (Gadfly release 3)
CG1401	BE978880, LP09387, LP03519, SD05659, LD34361, SD19264, RE35181, GH15159, RE32560, RE25661, LD15127, LD23103, RE55959
Analysis	CDNA
CG7042	AA203008 (491 bp mRNA, 2001), AW941523 (359 bp mRNA, 2001) (Flybase)
astray	AA820172 (581 bp mRNA, 2001), AF191498 (1466 bp mRNA, 2000; protein:AAF14696), AI455353 (601 bp mRNA, 2001), AY051689 (1524 bp mRNA, 2001; protein:AAK93113) (Flybase)
string	AI124307 (777 bp mRNA, 1999), AI515671 (676 bp mRNA, 2001), AW943922 (553 bp mRNA, 2001), AY069704 (2745 bp mRNA, 2001; protein:AAL39849), M24909 (2280 bp mRNA, 1993; protein:AAA28916), X57495 (2615 bp mRNA, 1992; protein:CAA40732) (Flybase)
CG1401	AY071504 (3524 bp mRNA, 2001; protein:AAL49126), BI369693 (555 bp mRNA, 2001) (Flybase)
Analysis	genomic DNA
CG7042	AE003467 (298640 bp DNA, 2000; protein:AAF47341) (Flybase)
astray	AE003552 (286784 bp DNA, 2000; protein:AAF50274), AJ271817 (17111 bp DNA, 2001; protein:CAB72249) (Flybase)
string	AE003768 (242815 bp DNA, 2000; protein:AAF56885) (Flybase)
CG1401	AE003768 (242815 bp DNA, 2000; protein:AAF56852) (Flybase)
Analysis	NCBI locus ID
CG7042	64868, Dm Mkp, MAP kinase-specific phosphatase (Aliases: DMKP, D-mkp; RefSeq: NM_080276; Nucleotide: AA142050, AA142051, AF250380; Protein: NP_525015, AAF67187)

astray	39085, Dm aay, astray, 67B4 (Aliases: CG3705, 0423/14, CT12429; RefSeq: NM_079277; Nucleotide: AE003552, AF174664, AF174665, AJ271817, AA820172, AF191498, AI455353, AY051689; Protein: NP_524001, AAF50274, AAF14696, AAK93113 (all 270 aa), CAB72249 (46 aa))
string	43466, Dm stg, string, 99A5 (Aliases: 5473, Cdc25, SY3-4, cdc25, CG1395, CT3224, 0224/06, 0245/03, 0439/22, 0730/13, 0896/05, 0967/05, 0980/06, 1083/13, 1089/08, 1143/02, S(rux)3A, l(3)j1D3, l(3)j1E3, l(3)j3D1, l(3)01235, l(3)j10B9, l(3)s2213, Cdc25[stg], clone 2.21, Cdc25[String], Cdc25[string], cdc25[string], anon-EST:Liang-2.21; RefSeq: NM_079823; Nucleotide: AE003768, AF174661, AF174662, AF174663, AQ025232, AQ073863, AQ074002, AQ074005, G00587, G00593, AI124307, AI515671, AW943922, AY069704, M24909, X57495; Protein: NP_524547, AAF56885, AAL39849, AAA28916, CAA40732 (all 479 aa))
CG1401	43434, Dm CG1401, 98F6 (Aliases: CT3252; RefSeq: NM_143408; Nucleotide: AE003768, AY071504; Protein: NP_651665, AAF56852, AAL49126)
Analysis	Drosophila mutations & mutants
CG7042	There are no recorded mutant alleles (Flybase)
astray	There is one recorded mutant allele, and it is available from the public stock centers
string	There are 97 recorded alleles: 16 in vitro constructs (1 available from the public stock centers), 80 classical mutants (3 available from the public stock centers) and 1 wild-type
CG1401	not described (Flybase)
Analysis	Phenotypic info
CG7042	not described (Flybase)
astray	Mutations have been isolated which affect the embryonic anterior fascicle and are pharate adult recessive lethal (Flybase)
string	Amorphic mutations have been isolated which affect the multidendritic neuron, the external sensory organ, the chordotonal organ and 10 other listed tissues and are embryonic recessive lethal, female fertile, recessive mitotic and somatic clone increased cell size (Flybase)
CG1401	not described (Flybase)

Example 3: Identification of human CG7042, astray, string, and CG1401 genes and proteins

CG7042, astray, string, and CG1401 homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are nucleic acids comprising Drosophila CG7042, astray, string, and CG1401, or human CG7042, astray, string, and CG1401 homologs (in particular the human protein phosphatase SKRP1, human phosphoserine phosphatase PSPH, cell division cycle 25A protein (CDC25A), cell division cycle 25B protein (CDC25B), cell division cycle 25C protein (CDC25C), and the human cullin 5 (CUL5)).

Sequences homologous to Drosophila CG7042, astray, string, and CG1401 were identified using the publicly available program BLASTP 2.2.3 of the non-redundant protein data base of the National Center for Biotechnology Information (NCBI)(see, Altschul et al., 1997, Nucleic Acids Res. 25:3389-3402). Table 2 shows the best human homologs of the Drosophila CG7042, astray, string, and CG1401.

Table 2. Human homolog proteins to Drosophila CG7042, astray, string, and CG1401 proteins

I. CG7042

* NCBI (National Center for Biotechnology Information) human locus identification (ID): 142679, Hs SKRP1, protein phosphatase, 2q32.1
* RefSeq[R]: GenBank Accession Number NM_080876 (58% homology of amino acids 13-200 of the Drosophila CG7042-PA protein to amino acids 15-198 of the human protein phosphatase SKRP1 (217 amino acids in total))
* Nucleotide: GenBank Accession Numbers AB038770, AB063186, AB063187

* Protein: GenBank Accession Numbers NP_543152 (217 amino acids), BAB82499 (217 amino acids), BAB83498 (217 amino acids), BAB83499 (166 amino acids)

* * * * Patents: GenBank Accession Number NP_543152 shows

* 100% identity to CAD10217.1, human unnamed protein product, disclosed in WO 0173060-A (Millennium Pharmaceuticals, Inc.),

* 100% identity to AX086030.1, human Sequence 27 from patent application WO0112819 (Sugen, Inc.),

* 100% identity to AX260334.1, human Sequence 1 from patent application WO0173060 (Millennium Pharmaceuticals, Inc.),

* 100% identity to AX287087.1, human Sequence 7 from patent application WO0181590 (Incyte Genomics, Inc.),

* 100% identity to AX260336.1, human Sequence 3 from patent application WO0173060 (Millennium Pharmaceuticals, Inc.)

*

II. astray

* NCBI (National Center for Biotechnology Information) human locus identification (ID): 5723, Hs PSPH, phosphoserine phosphatase, 7p15.2-p15.1

* Aliases: PSP

* OMIM: 172480

* RefSeq[R]: GenBank Accession Number NM_004577 (69% homology of amino acids 56-270 of Drosophila astray (GadFly Accession Number CG3705-PA) to amino acids 9-222 of human phosphoserine phosphatase (225 amino acids total))

* Nucleotide: GenBank Accession Number Y10275

* Protein: GenBank Accession Numbers NP_004568, CAA71318 (all 225 amino acids)* * * * * *

III. string

* NCBI (National Center for Biotechnology Information) human locus identification (ID): 993, Hs CDC25A, cell division cycle 25A, 3p21

* OMIM: 116947

* RefSeq[R]: GenBank Accession Number NM_001789 (50% homology of amino acids 116-461 of Drosophila string to amino acids 184-510 of human CDC25A (523 amino acids total))

* Nucleotide: GenBank Accession Numbers AF112978, AJ242714, BC007401, BC018642, M81933

* Protein: GenBank Accession Numbers NP_001780 (523 amino acids), AAH07401 (524 amino acids), AAH18642 (524 amino acids), AAA58415 (523 amino acids)

* NCBI (National Center for Biotechnology Information) human locus identification (ID): 994, Hs CDC25B, cell division cycle 25B, 20p13

* OMIM: 116949

* RefSeq[R]: GenBank Accession Numbers NM_004358 (52% homology of amino acids 63-461 of Drosophila string to amino acids 212-553 of human CDC25B (566 amino acids total)), NM_021872 (52% homology of amino acids 63-461 of Drosophila string to amino acids 185-526 of human CDC25B (539 amino acids total)), NM_021873 (52% homology of amino acids 63-461 of Drosophila string to amino acids 226-567 of human CDC25B (580 amino acids total)), NM_021874 (52% homology of amino acids 63-461 of Drosophila string to amino acids 247-588 of human CDC25B (601 amino acids total))

* Nucleotide: GenBank Accession Numbers AF036233, AL109804, X96436, BC006395, BC009953, M81934, S78187, Z68092

* Protein: GenBank Accession Numbers NP_004349 (566 amino acids), NP_068658 (539 amino acids), NP_068659 (580 amino acids), NP_068660 (601 amino acids), AAB94622 (297 amino acids), AAB94623 (283 amino acids), AAB94624 (256 amino acids), AAB94625 (305 amino acids), CAA65303 (62 amino acids), AAH06395 (580 amino acids), AAH09953 (580 amino acids), AAA58416 (566 amino acids), AAB21139 (566 amino acids), CAA92108 (539 amino acids)

* NCBI (National Center for Biotechnology Information) human locus identification (ID): 995, Hs CDC25C, cell division cycle 25C, 5q31

* Aliases: CDC25

* OMIM: 157680

* RefSeq[R]: GenBank Accession Numbers NM_001790 (63% homology of amino acids 251-461 of Drosophila string to amino acids 258-457 of human CDC25C (473 amino acids total)), NM_022809 (63% homology of amino acids 251-461 of Drosophila string to amino acids 185-384 of human phosphoserine phosphatase (400 amino acids total))

* Nucleotide: GenBank Accession Numbers Z29077, AF086323, AF277723, AF277724, AF277725, AF277726, AJ304504, BC019089, M34065

* Protein: GenBank Accession Number NP_001781 (473 amino acids), NP_073720 (400 amino acids), AAG41885 (149 amino acids), AAG41886 (90 amino acids), AAG41887 (136 amino acids), AAG41888 (106 amino acids), CAC19192 (400 amino acids), AAH19089 (473 amino acids), AAA35666 (473 amino acids)

IV. CG1401

* NCBI (National Center for Biotechnology Information) human locus identification (ID): 8065, Hs CUL5, cullin 5, 11q22-q23

* Aliases: VACM1, VACM-1

* OMIM: 601741

* RefSeq: GenBank Accession Number NM_003478 (81% homology of amino acids 202-852 of Drosophila CG1401 to amino acids 124-780 of human VACM-1 (780 amino acids total))

* Nucleotide: GenBank Accession Numbers AF017061, X81882

* Protein: GenBank Accession Numbers NP_003469 (780 amino acids), AAB70253 (781 amino acids), CAA57465 (780 amino acids)

* Patents: GenBank Accession Number NM_003478 shows

* 100% identity of amino acids 1-780 of the human NM_003478 to amino acids 82-861 of AAB47601, CUL5 (861 amino acids total), disclosed in WO0175145-A2 (RIGEL PHARM INC),

The mouse homologous cDNAs encoding the polypeptides of the invention were identified as GenBank Accession Numbers NM_024438 (for the mouse homolog to CG7042; Mm dual specificity phosphatase 19), NM_133900 (for the mouse homolog to astray; Mm expressed sequence AI480570), NM_007658 (for the mouse homolog to string; Mm cell division cycle 25 homolog A, Cdc25a), NM_023117 (for the mouse homolog to string; Mm cell division cycle 25 homolog B, Cdc25b), NM_009860 (for the mouse homolog to string; Mm cell division cycle 25 homolog C, Cdc25c), XM_134805 (for the mouse homolog to CG1401; Mm RIKEN cDNA 4921514I20 gene).

Claims

1. A pharmaceutical composition comprising a nucleic acid molecule of the CG7042 (protein phosphatase), the astray (phosphoserine phosphatase), the string (cell division cycle 25), or the CG61401 (cullin) gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an effector of said nucleic acid molecule or polypeptide together with pharmaceutically acceptable carriers, diluents and/or adjuvants.

2. The composition of claim 1, wherein the nucleic acid molecule is a vertebrate or insect CG7042, astray, string, or CG1401 nucleic acid, particulary encoding the human CG7042, astray, string, or CG1401 homologs, and/or a nucleic molecule which is complementary thereto or a fragment thereof or a variant thereof.

3. The composition of claim 1 or 2, wherein said nucleic acid molecule
   (a) hybridizes at 50°C in a solution containing 1 x SSC and 0.1% SDS to a nucleic acid molecule as defined in claim 2 and/or a nucleic acid molecule which is complementary thereto;
   (b) it is degenerate with respect to the nucleic acid molecule of (a),
   (c) encodes a polypeptide which is at least 85%, preferably at least 90%, more preferably at least 95%, more preferably at least 98% and up to 99,6% identical to the human CG7042, astray, string, or CG1401, as defined in claim 2;
   (d) differs from the nucleic acid molecule of (a) to (c) by mutation and wherein said mutation causes an alteration, deletion, duplication or premature stop in the encoded polypeptide.

4. The composition of any one of claims 1-3, wherein the nucleic acid molecule is a DNA molecule, particularly a cDNA or a genomic DNA.

5. The composition of any one of claims 1-4, wherein said nucleic acid encodes a polypeptide contributing to regulating the energy homeostasis and/or the metabolism of triglycerides.

6. The composition of any one of claims 1-5, wherein said nucleic acid molecule is a recombinant nucleic acid molecule.

7. The composition of any one of claims 1-6, wherein the nucleic acid molecule is a vector, particularly an expression vector.

8. The composition of any one of claims 1-5, wherein the polypeptide is a recombinant polypeptide.

9. The composition of claim 8, wherein said recombinant polypeptide is a fusion polypeptide.

10. The composition of any one of claims 1-7, wherein said nucleic acid molecule is selected from hybridization probes, primers and anti-sense oligonucleotides.

11. The composition of any one of claims 1-10 which is a diagnostic composition.

12. The composition of any one of claims 1-10 which is a therapeutic composition.

13. The composition of any one of claims 1-12 for the manufacture of an agent for detecting and/or verifying, for the treatment, alleviation and/or prevention of an disorders, including metabolic diseases such as obesity and other body-weight regulation disorders as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea and others, in cells, cell masses, organs and/or subjects.

14. Use of a nucleic acid molecule of the CG7042 (protein phosphatase), the astray (phosphoserine phosphatase), the string (cell division cycle 25), or the CG1401 (cullin) gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an effector of said nucleic or polypeptide for controlling the function of a gene and/or a gene product which is influenced and/or modified by a CG7042, astray, string, or CG1401 homologous polypeptide.

15. Use of the nucleic acid molecule of the CG7042 (protein phosphatase), the astray (phosphoserine phosphatase), the string (cell division cycle 25), or the CG1401 (cullin) gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an effector of said nucleic acid molecule or polypeptide for identifying substances capable of interacting with a CG7042, astray, string, or CG1401 homologous polypeptide.

16. A non-human transgenic animal exhibiting a modified expression of a CG7042, astray, string, or CG1401 homologous polypeptide.

17. The animal of claim 16, wherein the expression of the CG7042, astray, string, or CG1401 homologous polypeptide is increased and/or reduced.

18. A recombinant host cell exhibiting a modified expression of a CG7042, astray, string, or CG1401 homologous polypeptide.

19. The cell of claim 18 which is a human cell.

20. A method of identifying a (poly)peptide involved in the regulation of energy homeostasis and/or metabolism of triglycerides in a mammal comprising the steps of

(a) contacting a collection of (poly)peptides with a CG7042, astray, string, or CG1401 homologous polypeptide or a fragment thereof under conditions that allow binding of said (poly)peptides;

(b) removing (poly)peptides which do not bind and

(c) identifying (poly)peptides that bind to said CG7042, astray, string, or CG1401 homologous polypeptide.

21. A method of screening for an agent which modulates the interaction of a CG7042, astray, string, or CG1401 homologous polypeptide with a binding target/agent, comprising the steps of

(a) incubating a mixture comprising

(aa) a CG7042, astray, string, or CG1401 homologous polypeptide or a fragment thereof;

(ab) a binding target/agent of said CG7042, astray, string, or CG1401 homologous polypeptide or fragment thereof; and

(ac) a candidate agent

under conditions whereby said CG7042, astray, string, or CG1401 polypeptide or fragment thereof specifically binds to said binding target/agent at a reference affinity;

(b) detecting the binding affinity of said CG7042, astray, string, or CG1401 polypeptide or fragment thereof to said binding target to determine an (candidate) agent-biased affinity; and

(c) determining a difference between (candidate) agent-biased affinity and the reference affinity.

22. A method of producing a composition comprising the (poly)peptide identified by the method of claim 20 or the agent identified by the method of claim 21 with a pharmaceutically acceptable carrier, diluent and/or adjuvant.

23. The method of claim 22 wherein said composition is a pharmaceutical composition for preventing, alleviating or treating of diseases and disorders, including metabolic diseases such as obesity and other body-weight regulation disorders as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea and other diseases and disorders.

24. Use of a (poly)peptide as identified by the method of claim 20 or of an agent as identified by the method of claim 21 for the preparation of a pharmaceutical composition for the treatment, alleviation and/or prevention of of diseases and disorders, including metabolic diseases such as obesity and other body-weight regulation disorders as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea and other diseases and disorders.

25. Use of a nucleic acid molecule of the CG7042 (protein phosphatase), the astray (phosphoserine phosphatase), the string (cell division cycle 25), or the CG1401 (cullin) gene family or of a fragment thereof for the preparation of a non-human animal which over- or under-expresses the CG7042, astray, string, or CG1401 gene product.

26. Kit comprising at least one of

(a) a CG7042, astray, string, or CG1401 nucleic acid molecule or a fragment thereof;

(b) a vector comprising the nucleic acid of (a);

(c) a host cell comprising the nucleic acid of (a) or the vector of (b);

(d) a polypeptide encoded by the nucleic acid of (a);

(e) a fusion polypeptide encoded by the nucleic acid of (a);

(f) an antibody, an aptamer or another receptor against the nucleic acid of (a) or the polypeptide of (d) or (e) and

(g) an anti-sense oligonucleotide of the nucleic acid of (a).

Abstract

The present invention discloses CG7042, astray, string, or CG1401 homologous proteins regulating the energy homeostasis and the metabolism of triglycerides, and polynucleotides, which identify and encode the proteins disclosed in this invention. The invention also relates to the use of these sequences in the diagnosis, study, prevention, and treatment of diseases and disorders, for example, but not limited to, metabolic disorders and diseases such as the metabolic syndrome, including obesity, eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancers of the reproductive organs, and sleep apnea.

ld 27.09.2002

FIGURE 1. Triglyceride content of a Drosophila CG7042 (GadFly Accession Number CG7042-PA) mutant

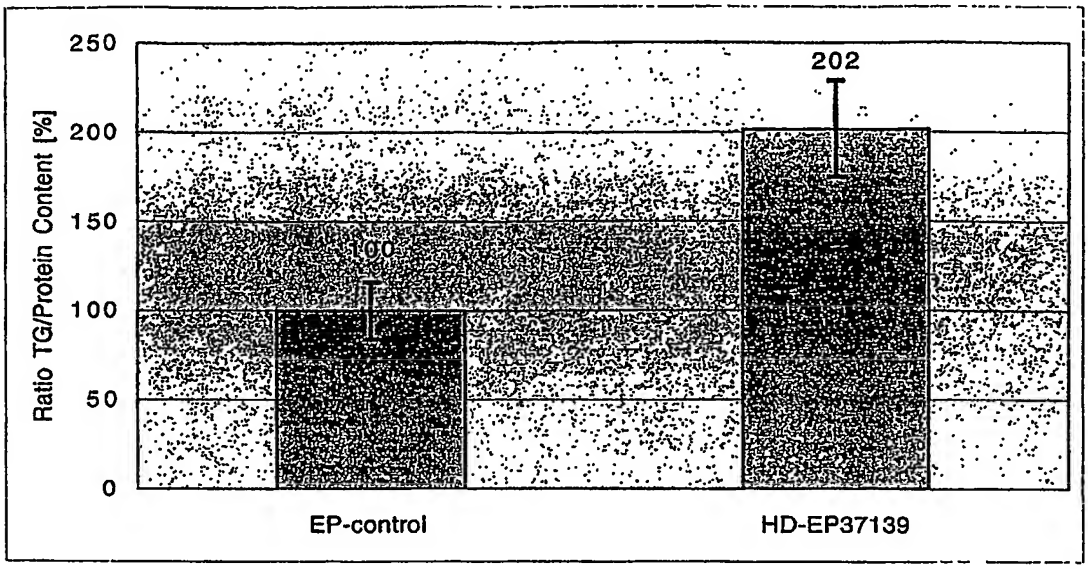

FIGURE 2. Molecular organization of the CG7042 gene (GadFly Accession Number CG7042-PA)

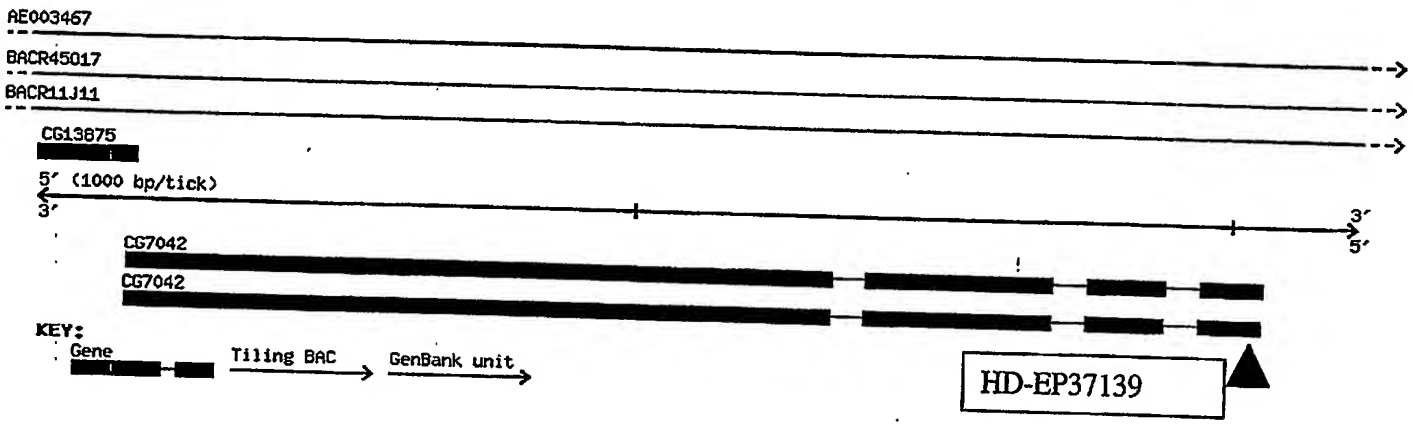

FIGURE 3. Triglyceride content of a Drosophila astray (GadFly Accession Number CG3705) mutant

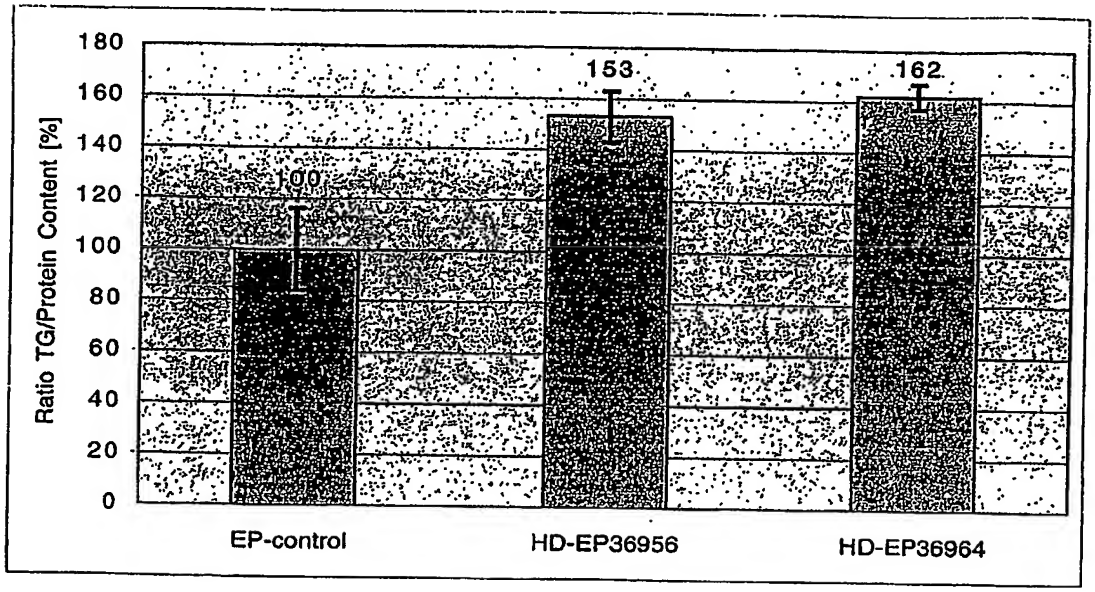

FIGURE 4. Molecular organization of the astray gene (GadFly Accession Number CG3705)

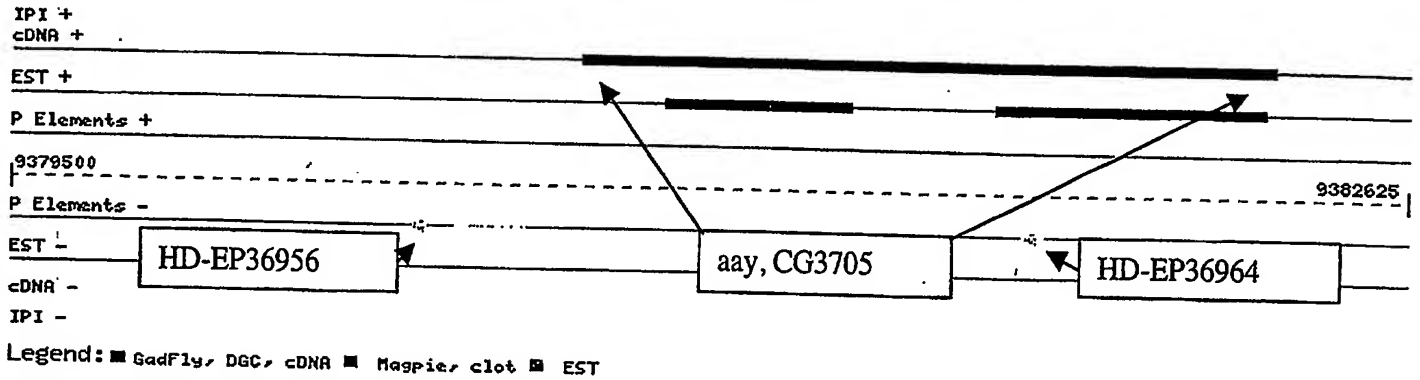

FIGURE 5. Triglyceride content of a Drosophila string (GadFly Accession Number CG1395) mutant

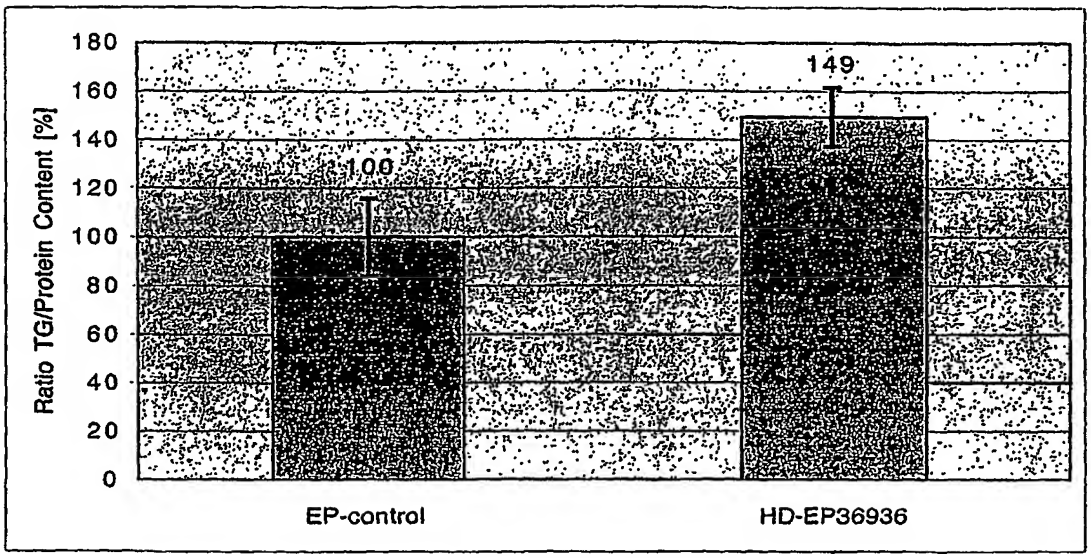

FIGURE 6. Molecular organization of the string gene (GadFly Accession Number CG1395)

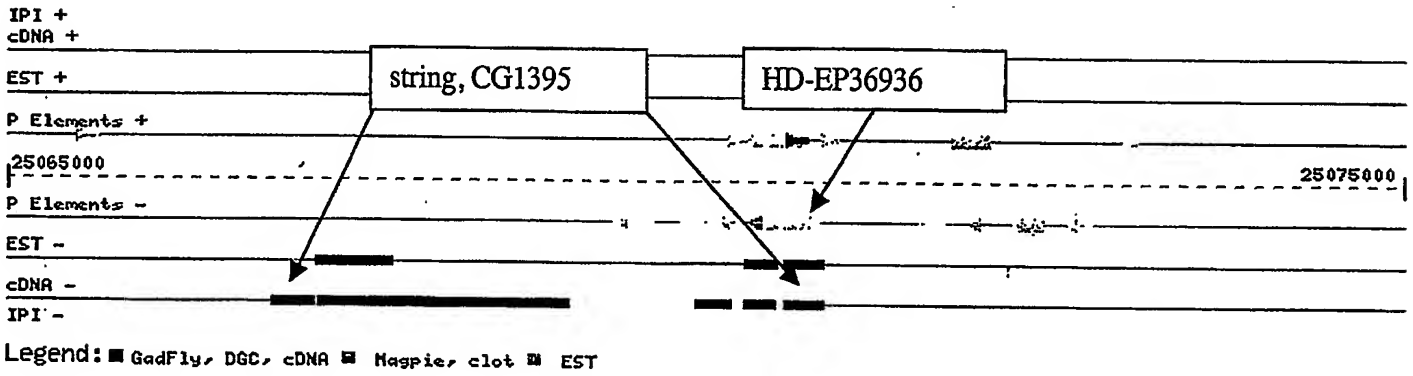

FIGURE 7. Triglyceride content of a Drosophila CG1401 (GadFly Accession Number) mutant

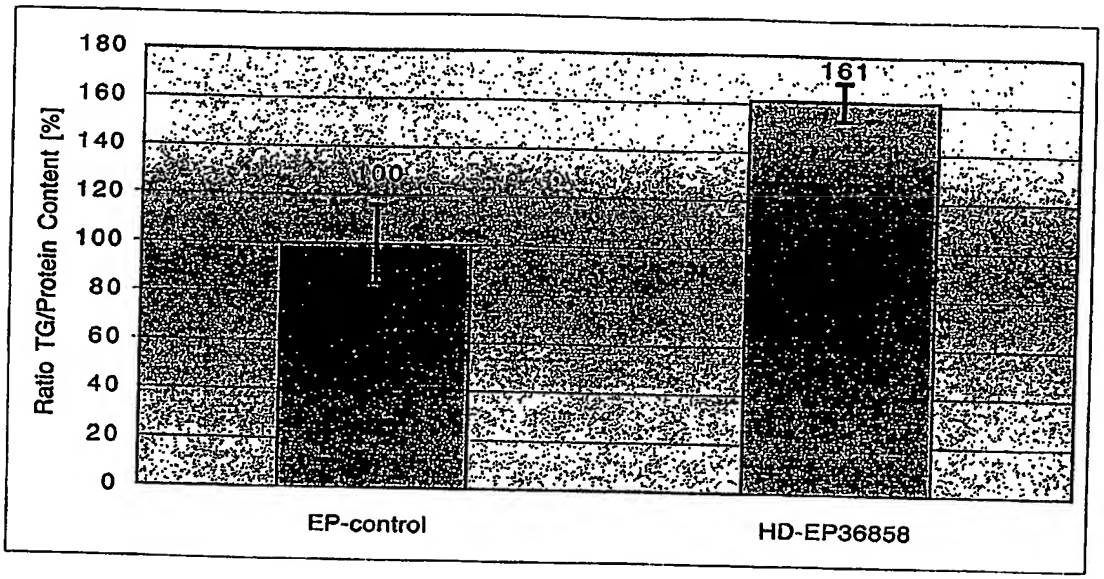

FIGURE 8. Molecular organization of the CG1401 gene (GadFly Accession Number)